VILNIAUS UNIVERSITETAS

TEORINĖS FIZIKOS KATEDRA

Thomas Gajdosik

Special Relativity for Particle Physics

notes for the lecture Įvadas į elementariųjų dalelių fiziką

Vilnius 2013

Leidinį T. Gajdosik. ”Special Relativity for Particle Physics” apsvarstė ir rekomendavo spaudai VU teorinės fizikos katedra (2012-06-26, protokolas Nr. 07-2012).

Recenzavo: dr. Artūras Acus

Prof. dr. (HP) Egidijus Anisimovas

Contents

1 Introduction **4**

2 Space and Time **4**
2.1 Representing Space and Time . . . 4
2.2 Vectors and Vector Space . . . 5
2.2.1 Examples . . . 7

3 Changes between coordinate systems **8**
3.1 Translations . . . 9
3.2 Rotations . . . 10

4 General index notation **12**
4.1 Tensor product . . . 13
4.2 Special tensors: the Kronecker delta . . . 14
4.3 Symmetries of a tensor . . . 14
4.4 Special tensors: the Levi-Civita symbol . . . 15
4.5 Restrictions on coordinate transformations . . . 15
4.5.1 Remark on the shortcomings of this introduction . . . 16

5 Invariants **16**

6 Galilean transformations **18**
6.1 The pendulum under Galilean transformations . . . 19

7 Four dimensional spacetime **22**
7.1 Four-vectors . . . 23
7.2 The metric . . . 23
7.3 Length measurement . . . 25
7.4 Time measurement . . . 25
7.5 Lorentz transformations . . . 26
7.6 Four-momentum . . . 29

References **31**

Acknowledgments

I want to thank dr. Artūras Acus for reading the manuscript very carefully, for his help in editing the text and for giving valuable suggestions.

1 Introduction

The primary audience for this notes are students of the course *World of Particles* (Įvadas į elementariųjų dalelių fiziką). These notes are written mainly for beginning students that have difficulty imagining, how a mathematical formulation can be connected to a physical content. Therefore I will pay more attention to explain some concepts exhaustively while missing other things that seem too obvious to me — or that I think are explained well enough in the freely available text of David Hogg [1]. Corrections and feedback are welcome as this text is written to help students.

2 Space and Time

Physics describes the world around us, but seen through the scientific method: the Physics viewpoint abstracts from the individual perceptions to capture what is common and repeatable in the everyday experience of human beings. For this purpose physics uses the universal language of mathematics.

One of the basic features of our world is that time passes continuously and that our perceptions can be most accurately and economically described by a three dimensional space. The discussion of the continuity of the three dimensional space was studied in geometry, starting with Euclid and Archimedes. René Descartes connected geometry with algebra by introducing the Cartesian coordinates to describe relations in space. The introduction of the differential calculus into physics by Newton relies heavily on the use of Cartesian coordinates. The continuity of time, measured by clocks, suggests the use of continuous mathematical functions to describe changes in our world. Even the very abrupt changes, like explosions, can be described by continuous functions of time.

2.1 Representing Space and Time

When we try to mathematically describe a motion, like the examples given in the Problems 1-1 and 1-2 of [1], we "instinctively" use Cartesian coordinates for space and time: we give distances in meters and time in hours (or seconds) and the meter does not depend on time and the hour not on the distance. This is the virtue of Cartesian coordinates: they are independent from each other.

Considering the three dimensions of space we can introduce also three independent directions, which make up a Cartesian coordinate system. The usual names for these axes are x, y, and z. Then we can describe the position of an object in our coordinate system by giving its values along the coordinate axes. Assigning the x-axis to my direction of sight, the x coordinate tells how far in front of me the object is, the y coordinate tells how far to the left of my line of sight the object is, and the z coordinate tells how high the object is with respect to me[1]. A negative value of the coordinate indicates that the object is behind/right/under me. Since I give the values with respect to my position, I implicitly introduced the origin of the coordinate system O and identified it with myself.

[1] This choice of directions is called a right-handed coordinate system. A left-handed coordinate system has the direction of the y-axis reversed.

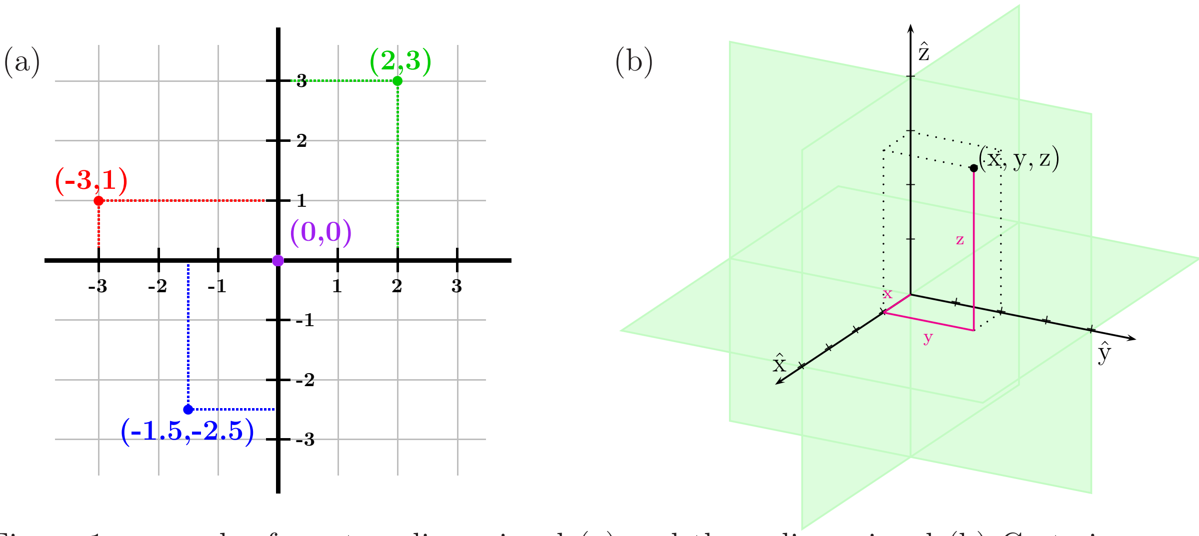

Figure 1: examples for a two dimensional (a) and three dimensional (b) Cartesian coordinate system. In (a) are the origin $(0,0)$ in purple and the points $(-3,1)$ in red, $(2,3)$ in green, and $(-1.5,-2.5)$ in blue. In (b) is the single point (x,y,z) $= (1,2,3)$ with its coordinate parts in magenta.

We can introduce the vector pointing from me to an object A by writing down its coordinates in triple of numbers: $\vec{a} := (a_x, a_y, a_z)$. Of course, the vector from me to the object A makes sense, even if I do not introduce a coordinate system. This viewpoint allows the idea of a vector space. [2] takes vectors in their coordinate independence even more seriously and proposes an approach that could unify the mathematical description in different areas of physics.

Fig.1(a) gives an example of a Cartesian coordinate system in two dimensions, which amounts to ignore the z coordinate. Fig.1(b) gives an example of a Cartesian coordinate system in three dimensions.

2.2 Vectors and Vector Space

The direct definition of a vector is mathematically quite difficult. Mathematicians solve this problem by making two steps:

Definition: A vector is an element of a vector space.

Definition: A vector space over a field (corpus) F is a set V with two operations "+" and "$*$" such that $(V,+)$ forms an Abelian group and $(F,+,*)$ forms a field (corpus).

Now what does that mean? Group Theory describes what a set, a group, and a field (corpus) is. For the mathematically more dedicated physics student I suggest the lecture notes of Minahan [3]. For here the simple introduction should suffice.

First we have to clarify what a set is. A set is an assembly of things, of what ever you can think of. You just have to declare, that they belong to the set. Of course, one can also give a descriptive definition for a specific set. The easiest example might be from mathematics: the set of the numbers $\{0, 1, 2\}$ can also be described as the remainder of the division of whole numbers by three.

A group has already more structure, the group "multiplication". This does not need to be a multiplication we know from numbers. In the above example, when taking the numbers $\{0, 1, 2\}$ we can use the normal addition of numbers as the group multiplication. We just have to remember, that after each addition we have to divide by three and only take the remainder of this division. We thereby also include the descriptive definition for our set. This set G (i.e. $\{0, 1, 2\}$) together with the addition and taking the remainder (i.e. $\circ$) forms a group as it fulfills these axioms:

closure: for $a, b \in G \Rightarrow c = a \circ b \in G$
: the group multiplication connects two elements of the set and the result is still an element of the set.

distributivity: $(a \circ b) \circ c = a \circ (b \circ c)$
: when multiplying three elements it does not matter if we first multiply the first pair and then the result with the third element or if we first multiply the second pair and only afterward multiply the first element with the result. But we are not allowed to change the order of the elements.

unit element: $\forall a \in G : \exists e \in G$ with $a \circ e = e \circ a = a$
: this unit element is easier to understand, when we consider that elements of a set can also be abstract things like transformations. The multiplication with a group element a means that I transform the other element b, which I multiply with a: $a \circ b$ is a left "action" on b (I multiply from the left side) and $b \circ a$ is a right action, as I multiply b with a from the right side. The unit element e now means, that I do not transform at all: I transform with the identity element. And each group has to have an identity element.

inverse element: $\forall a \in G : \exists a^{-1} \in G$ with $a \circ a^{-1} = a^{-1} \circ a = e$
: when understanding the inverse element as a transformation this axiom just means, that we can always transform back to the original state. For the realization with numbers we see immediately, that this axiom really restricts, what we can name a group. The real numbers, for instance, with the normal multiplication are not a group, as there is no possibility to find an inverse element for the number zero.

: In the case of our example $\{0, 1, 2\}$, the inverse element is also not too obvious. The unit element is simply 0, as we do not change a number, if we add 0. We can also see, that when we add $1 + 2$ we get 3, which is equivalent to 0, as there is no remainder when we divide by three. So 1 is the inverse element to 2 and vice versa. That 0 is the inverse element of itself seems counter intuitive, but when we consider that 0 is the unit element, it becomes easier to accept, that multiplying the unit element with itself has to give the unit element, and hence it has to be its own inverse.

There is another concept connected with groups. When we multiply two elements we can do it in either order: $a \circ b$ or $b \circ a$. If the result is the same for all elements of the group, i.e. $a \circ b = b \circ a \ \forall a, b \in G :$, the group is called Abelian. If that is not the case, if only a single pair of elements (a, b) exists, so that $a \circ b \neq b \circ a$, then the group is non-Abelian.

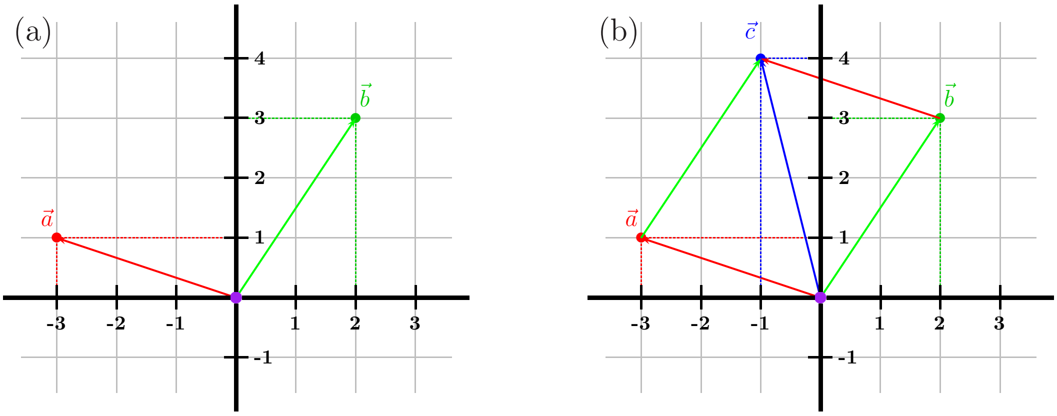

Figure 2: an example for two dimensional vectors (a) and their addition (b). The red vector $\vec{a}$ has the x component -3 and the y component 1. The green vector $\vec{b}$ has the x component 2 and the y component 3. When we add $\vec{a}+\vec{b}=\vec{c}$ we get the x component of the blue vector $\vec{c}$ as the sum of the x components of the vectors $\vec{a}$ and $\vec{b}$, which is -1.

Now with the understanding of an Abelian group we can follow the definition of a field (corpus)[2]. A corpus F is like a double Abelian group. It has two group operations, called plus "+" and times "*". And regarding the plus the corpus $(F,+)$ is an Abelian group. Its unit element is called "0". Regarding the times we have to remove this unit element from the set and then the rest, mathematically $(F\backslash\{0\},*)$, is again an Abelian group. In that sense the corpus is a double Abelian group. Additionally, the two operations of the corpus have to be distributive: $a*(b+c)=a*b+a*c$.

We are looking to understand vectors and vector spaces: so enough about groups, but we will need to use the concepts. The vector space has the plus operation that lets one add vectors, the elements of the Abelian group, that the vector space is. This plus is the same as the plus of the corpus. This will allow us to recognize the components of vectors as numbers, the elements of the corpus. The vector space has to be distributive in the multiplication of numbers with the vectors. With vectors $\vec{x}$ and $\vec{y}$ and numbers a and b we can write the distributive relations as

$$a*(\vec{x}+\vec{y}) = a*\vec{x}+a*\vec{y} \tag{1}$$
$$(a+b)*\vec{x} = a*\vec{x}+b*\vec{x} \tag{2}$$
$$a*(b*\vec{x}) = (a*b)*\vec{x}\ . \tag{3}$$

2.2.1 Examples

The first example of a vector and a vector space will be what everybody expects: the real, two dimensional vectors in a plane. Formally we can call this vector space $\mathbf{R}^2$ over $\mathbf{R}$: the vectors, as shown in Fig. 2, can be written just with their components which are real numbers. Adding two vectors gives again a vector and we can get its components by adding the components of the vectors we want to add.

[2]From now on I will drop the mathematical name field and use just the mathematical term corpus instead, as the name field will be used to describe functions of space and time, like the electromagnetic field.

The second example of a vector space are the real numbers. Since the real numbers satisfy all the requirements for a vector space, they can be treated as a vector space. That means, that a real number can be understood as a vector, too.

The third example are matrices of a fixed dimension. We can add the matrices by adding the elements of the matrices in the same way as we add the components of vectors. The structure of the vector space does not use matrix multiplication. This would be an additional feature, like matrices forming a group under multiplication.

The fourth example are real analytic functions in the interval from zero to one: we can add them and multiply them with numbers and they still stay real analytic functions. The huge difference between this example and the previous ones is the dimensionality of the vector space. In the first example we can think of the two vectors $\vec{a}$ and $\vec{b}$ as generating the whole plane: I can reach any point in the plane by adding real multiples of the two vectors. In the second example I only have a single direction: each real number can be reached by multiplying the number 1 with the desired value. The third example has as many dimensions as the matrices have entries: the number of rows times the number of columns. But for the fourth example it is no longer easy to find a simple way of generating all possible real analytic functions. There are infinitely many possibilities.

3 Changes between coordinate systems

We used the notion of a coordinate system like a grid, where we can read off the coordinates of points. This grid is our tool. We are able to change the grid without changing the points and their differences. The description of this change is a coordinate transformation. The most general coordinate transformation gives the new coordinates (t', x', y', z') as functions of the old coordinates (t, x, y, z):

$$t'(t,x,y,z) \qquad x'(t,x,y,z) \qquad y'(t,x,y,z) \qquad z'(t,x,y,z) \ . \tag{4}$$

In order that this coordinate transformation makes sense, the functions have to be invertible. One has to be able to express the old coordinates in terms of the new ones:

$$t(t',x',y',z') \qquad x(t',x',y',z') \qquad y(t',x',y',z') \qquad z(t',x',y',z') \ . \tag{5}$$

The study of these most general coordinate transformations leads to the theory of General Relativity. But we do not want to go that far. We just want to understand, what these coordinate transformations imply for vectors.

The first step will be the comparison between two Cartesian coordinate systems. For that we have to ask, what defines a Cartesian coordinate systems. The most obvious definition is to require that the coordinate lines are straight lines that are orthogonal to each other. Then the grid made of these lines is what we see in Fig. 1 and Fig. 2. We have also to pick an origin of the coordinate system. This is the point where we start with our coordinates. So the origin itself has always the coordinates $(0, 0)$ in two dimensions, $(0, 0, 0)$ in three dimensions, and $(0, 0, 0, 0)$ if we also include time into our coordinates.

When we try to do this procedure for the surface of our Earth, we find out, that it does not work. We notice, that the surface of our Earth is not flat, but curved, which prevents the global use of a Cartesian coordinate system. But in small portions, like the

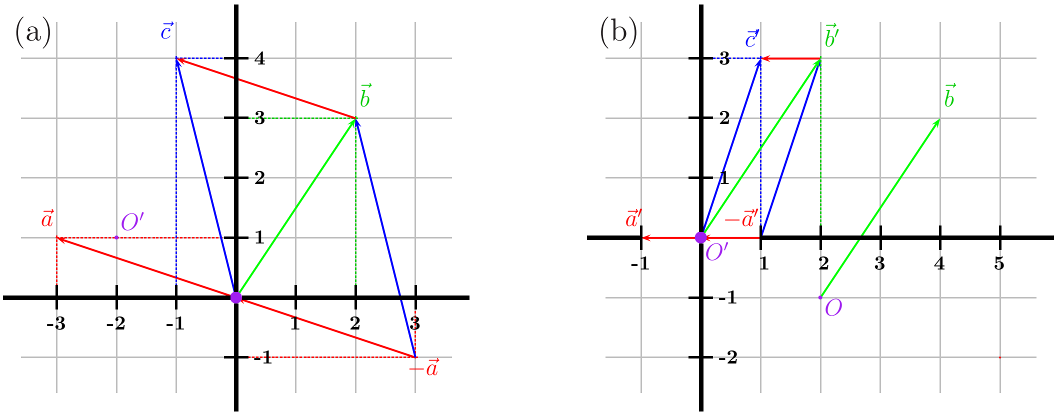

Figure 3: the difference of the vectors $\vec{c} - \vec{a} = \vec{b}$ in the original Cartesian coordinate system (a) and the new Cartesian coordinate system O' (b). In O, the origin of the new coordinate system O' has the coordinates (-2,1), as seen in (a). In (b) we see the origin of the old coordinate system O with coordinates (2,-1). Though the vector $\vec{b}$ appears in both coordinate systems at different locations, it has the same coordinates (2,3).

classroom it works without any problem. We call a space where Cartesian coordinates work Euclidean space, although this is not the mathematical definition of Euclidean space.

3.1 Translations

The easiest coordinate transformation is, when we have two Cartesian coordinates with a different origin, but the directions and the names of the coordinates are the same. Then the coordinates are related to each other by adding a constant value to the coordinates in one system. For simplicity the example will be in two dimensions. As we can see in Fig 3, the shift of the origin is the simplest coordinate transformation. We can write the transformation as

$$x' = x + 2 \qquad \text{and} \qquad y' = y - 1 \ . \tag{6}$$

This means we get for the vectors $\vec{a} = (-3, 1)$ and $\vec{c} = (-1, 4)$ the transformation

$$a'_x = a_x + 2 = -1 \qquad a'_y = a_y - 1 = 0 \quad \text{and} \quad c'_x = c_x + 2 = 1 \qquad c'_y = c_y - 1 = 3 \ , \tag{7}$$

or in row vector format

$$\vec{a}' = \vec{a} + (2, -1) = (-1, 0) \quad \text{and} \quad \vec{c}' = \vec{c} + (2, -1) = (1, 3) \ , \tag{8}$$

or in column vector format[3]

$$\vec{a}' = \vec{a} + \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \quad \text{and} \quad \vec{c}' = \vec{c} + \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \ . \tag{9}$$

[3]Both, row vectors and column vectors form vector spaces. But they are usually used to indicate *different* vector spaces: you cannot add a row vector to a column vector! In this example I can use the same components to describe both, row and column vectors, only because I use a Cartesian coordinate system.

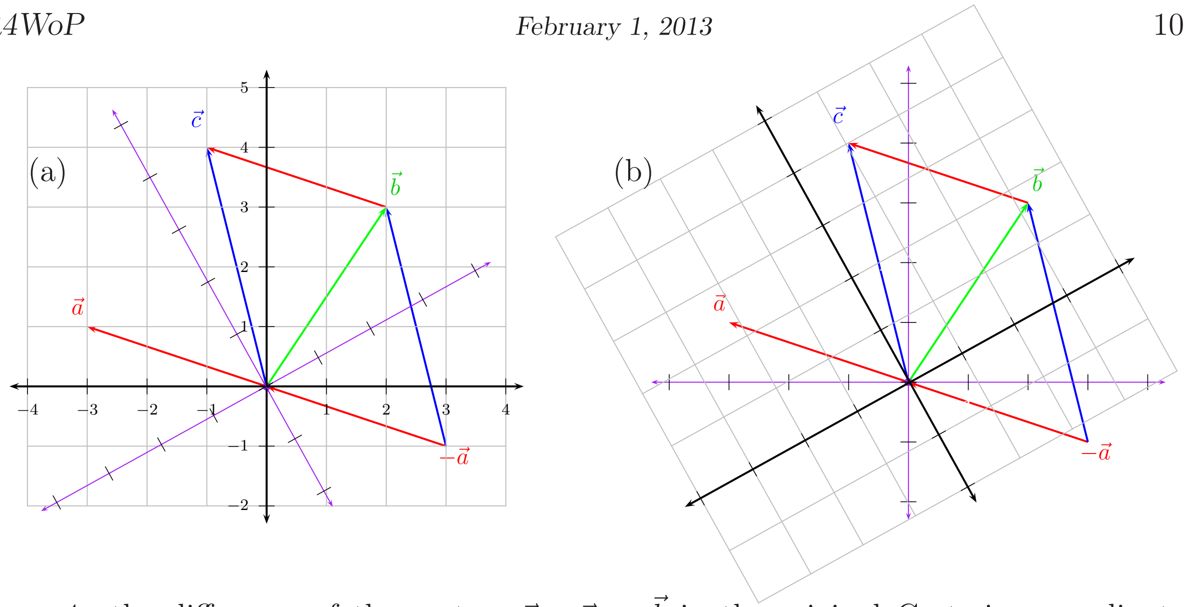

Figure 4: the difference of the vectors $\vec{c} - \vec{a} = \vec{b}$ in the original Cartesian coordinate system (a) and the rotated Cartesian coordinate system O' (b). The axes of the other coordinate system are shown in purple. The coordinates of the vectors in O are $\vec{a} = (-3, 1)$, $\vec{c} = (-1, 4)$, and $\vec{b} = (2, 3)$, and in O' they are $\vec{a}' = (-2.14, 2.33)$, $\vec{c} = (1.06, 3.98)$, and $\vec{b} = (3.2, 1.65)$.

Since the vector $\vec{b}$ in this example is defined as the difference between $\vec{c}$ and $\vec{a}$ its dependence on the origin drops out. This is the behavior that we expect from physical quantities: they should not depend on our choice of a coordinate system.

A shift of the origin can be called translation. That physics should be invariant under translations can be traced back to Galileo's principle of relativity. About that later ...

3.2 Rotations

The rotation between two Cartesian coordinate systems changes the coordinates of the points and the components of the vectors in the same way as can be seen in Fig. 4. We take the same points A, C, and B represented by the vectors $\vec{a}$, $\vec{c}$, and $\vec{b}$ from the origin to the respective point. These vectors are called position vectors or radius vectors, as the radius does not change with the rotation around the origin:

$$|\vec{a}| = \sqrt{(-3)^2 + 1^2} = \sqrt{10} \sim 3.16 \quad \text{and} \quad |\vec{a}'| = \sqrt{(-2.14)^2 + (2.33)^2} \sim 3.16 \tag{10a}$$
$$|\vec{c}| = \sqrt{(-1)^2 + 4^2} = \sqrt{17} \sim 4.12 \quad \text{and} \quad |\vec{c}'| = \sqrt{(1.06)^2 + (3.98)^2} \sim 4.12 \tag{10b}$$
$$|\vec{b}| = \sqrt{2^2 + 3^2} = \sqrt{13} \sim 3.61 \quad \text{and} \quad |\vec{b}'| = \sqrt{(3.2)^2 + (1.65)^2} \sim 3.61 \ . \tag{10c}$$

The coordinate transformation is no longer as simple as eq. (6), since with the rotation the change depends on the coordinates themselves. We know, that the distance to the origin stays the same in a rotation. Using this fact we can try a linear ansatz for the change of y- and x- coordinate:

$$x' = a_{11}x + a_{12}y \qquad \text{and} \qquad y' = a_{21}x + a_{22}y \ . \tag{11}$$

These two equations should hold for all three transformed vectors:

$$\begin{aligned}
\vec{a}:&\quad -2.14 = a_{11}\times(-3)+a_{12}\times 1 \quad\text{and}\quad 2.33 = a_{21}\times(-3)+a_{22}\times 1 &\text{(12a)}\\
\vec{c}:&\quad 1.06 = a_{11}\times(-1)+a_{12}\times 4 \quad\text{and}\quad 3.98 = a_{21}\times(-1)+a_{22}\times 4 &\text{(12b)}\\
\vec{b}:&\quad 3.2 = a_{11}\times 2+a_{12}\times 3 \quad\text{and}\quad 1.65 = a_{21}\times 2+a_{22}\times 3\ . &\text{(12c)}
\end{aligned}$$

We will solve these equations by elimination. Adding two times eq. (12b) to eq. (12c) we get

$$2\vec{c}+\vec{b}:\quad 2.12+3.2 = 11a_{12} \quad\text{and}\quad 7.96+1.65 = 11a_{22}\ , \tag{13}$$

so

$$a_{12} = \tfrac{5.32}{11} \sim 0.48 \quad\text{and}\quad a_{22} = \tfrac{9.61}{11} \sim 0.87\ , \tag{14}$$

and putting the result in eq. (12b)

$$a_{11} = 4\times 0.48 - 1.06 \sim 0.87 \quad\text{and}\quad a_{21} = 4\times 0.87 - 3.98 \sim -0.48\ . \tag{15}$$

Checking the result for eq. (12a)

$$-2.14 = -3\times 0.87 + 0.48 \sim -2.13 \quad\text{and}\quad 2.33 = -3\times(-0.48)+0.87 \sim 2.31\ , \tag{16}$$

gives the confirmation with the expected accuracy.

We can also write the rotation in matrix form[4]

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}\ , \tag{17}$$

which explains the indices of a in the ansatz eq. (11). The short vector form of eq. (17) is

$$\vec{x}' = \mathbf{R}\cdot\vec{x}\ , \tag{18}$$

with

$$\vec{x}' = \begin{pmatrix} x' \\ y' \end{pmatrix}\ , \qquad \mathbf{R} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\ , \text{and} \qquad \vec{x} = \begin{pmatrix} x \\ y \end{pmatrix}\ , \tag{19}$$

which can also be written in index form

$$x'_j = \sum_{k=1}^{2} \mathbf{R}_{jk}x_k := \mathbf{R}_{jk}x_k = a_{jk}x_k\ , \tag{20}$$

where the last definition is Einsteins summation convention: it tells you to sum over all indices that appear exactly two times in one term. The convention just allows to leave out the symbol $\sum$, which shortens the writing.

[4]Anyone not familiar with matrix multiplication can find an explanation in Wikipedia, searching for "Matrix multiplication".

4 General index notation

The introduction of the index notation shortens the writing of equations not only by omitting the summation symbol and reducing sums to apparent single terms, but by writing many equations in a single line. Writing eq. (11) as the last part of eq. (20) simplifies also the understanding of the different parts of the equation. Each object that has an index stands for all the parts that are summarized with the index: x_k of eq. (20) stands for x **and** y. That tells also, that x and y are of the same type and that x and y should transform in a similar way. The same holds for x'_j. For $a_{jk} = \mathbf{R}_{jk}$ it tells us, that all the terms a_{jk} belong to the single quantity $\mathbf{R}$, which we could write as a matrix, as seen in eq. (19). In principle, nothing prevents us from using more indices in a quantity. We cannot write this quantity as a matrix anymore, as a matrix has rows and columns, but not a generalization to more dimensions.

This feature of an indexed quantity allows another definition of the vector and the generalization to tensors. When we view the general coordinate transformations[5], eq. (4), we can restrict ourselves to a small region around the place we are interested in. Mathematically speaking, we select a point $\vec{x}_0$ and a neighborhood U of this point, where we discuss the coordinate transformations. Choosing the point $\vec{x}_0$ amounts to choosing a local origin for our coordinate transformations. Then we Taylor expand the coordinate transformation around the point $\vec{x}_0$ and keep only terms up to first order in the original coordinates:

$$x' = b_1 + a_{11}x + a_{12}y + a_{13}z \tag{21a}$$
$$y' = b_2 + a_{21}x + a_{22}y + a_{23}z \tag{21b}$$
$$z' = b_3 + a_{31}x + a_{32}y + a_{33}z \ . \tag{21c}$$

We can "justify" this linear approximation with the mathematical observation, that higher powers of a small quantity, i.e. something much smaller than 1, are smaller than the quantity itself. The smaller we choose the neighborhood U, the better the linear approximation will be. The quantities b_i in eq. (21) correspond to the translations we discussed in Sec. 3.1 and the quantities a_{jk} correspond to the rotations discussed in Sec. 3.2. The short index notation for eq. (21) is

$$x'_j = b_j + a_{jk}x_k \ , \tag{22}$$

which still are three equations, one equation for each index value of j. This linear set of equations is mathematically called an *affine* transformation.

One part of the Taylor expansion of the general coordinate transformations was that we had to pick the point around which we do the Taylor expansion. This amounts to choosing the origin of the old and the new coordinate system. When choosing the origin of both to be the reference point of the Taylor expansion, we can avoid to include the terms b_j in eq. (22), which is the same as setting $b_j = 0$, resulting in *homogeneous* transformations. Then the only thing that determines the transformation are the coefficients a_{jk}, which are

[5]For now we leave out the transformation of time. We will first introduce the common convention with upper and lower indices, which run from 0 to 3, with 0 meaning the time coordinate, before we discuss transformations that include time.

just the coefficients of the first order in the Taylor expansion. That means, we can write

$$a_{jk} = \frac{\partial x'_j}{\partial x_k} \ . \tag{23}$$

The transformation for the components of a vector $\vec{v} = v_j$, as can be seen implicitly from eq. (12), are

$$v'_j = a_{jk} v_k = \frac{\partial x'_j}{\partial x_k} v_k \ , \tag{24}$$

which can be used as a **definition**, that $\vec{v} = v_j$ is a vector. This type of definition for a vector can be found especially in older textbooks. It used the "natural" understanding of the coordinate systems to define the properties of a vector and to extend the definition to tensors. A newer compilation of this approach can be found in [4].

Definition: The components of a tensor transform under coordinate transformations in the same way as the coordinates themselves:

$$t'_{j_1 j_2 \ldots j_N} = \frac{\partial x'_{j_1}}{\partial x_{k_1}} \frac{\partial x'_{j_2}}{\partial x_{k_2}} \cdots \frac{\partial x'_{j_N}}{\partial x_{k_N}} t_{k_1 k_2 \ldots k_N} \ . \tag{25}$$

Every object that transforms in this way is a tensor. A vector is just a tensor with a single index. A matrix is a tensor with two indices.

4.1 Tensor product

According to the definition eq. (25) a quantity obtained from the multiplication of two tensors, is again a tensor:

$$t_{j_1 j_2 \ldots j_M k_1 k_2 \ldots k_N} = a_{j_1 j_2 \ldots j_M} \otimes b_{k_1 k_2 \ldots k_N} \ . \tag{26}$$

Of course, a product of two vectors (tensors with a single index) is not again a vector, because it has not one, but two indices. But we can understand this product as matrix:

$$a_j \otimes b_k = M_{jk} \tag{27}$$

or

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \otimes \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix} = \begin{pmatrix} a_1 b_1 & a_1 b_2 & a_1 b_3 \\ a_2 b_1 & a_2 b_2 & a_2 b_3 \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \end{pmatrix} \ , \tag{28}$$

where I have chosen to arrange the tensor M_{jk} with j being the index for the rows and k being the index for the columns. I could also have chosen differently:

$$\begin{pmatrix} a_1 & a_2 \end{pmatrix} \otimes \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_1 b_1 & a_2 b_1 \\ a_1 b_2 & a_2 b_2 \\ a_1 b_3 & a_2 b_3 \end{pmatrix} = \begin{pmatrix} M_{11} & M_{21} \\ M_{12} & M_{22} \\ M_{13} & M_{23} \end{pmatrix} \ , \tag{29}$$

which does not change anything in the way, M_{jk} is related to a_j and b_k. In this example eq. (27) contains six separate equations, a separate equation for each index pair jk. Only in the above case this **tensor product** (or *direct product* or *Cartesian product*) can be written in matrix form. When we have more indices the two dimensions of a matrix are no longer enough to give each index its own dimension.

4.2 Special tensors: the Kronecker delta

A special tensor with two indices is the **Kronecker delta** δ_{jk}. Both indices have to have the same index range, not like M_{jk} of the previous example, which has the index ranges for the first index $j = 1, 2$ and for the second index $k = 1, 2, 3$. The values of the components of the Kronecker delta are 0 or 1: only when both indices are equal, i.e. $j = k$, the value of $\delta_{j=k} = 1$. When both indices are not equal, i.e. $j \neq k$, the value of $\delta_{j\neq k} = 0$.

Using Einsteins summation convention and multiplying a tensor with the Kronecker delta, which has indices that appear on the tensor, is called contraction:

$$t_{jkn} \otimes \delta_{jk} = \sum_{j,k} t_{jkn}\delta_{jk} = \sum_{j} t_{jjn} = t'_n \quad . \tag{30}$$

Both indices j and k in t_{jkn} have to have the same index range. Tensor product and contraction together allow a definition of the usual matrix multiplication:

$$A \cdot B = C \tag{31}$$

with the $p \times r$-matrix A, the $r \times q$-matrix B, and the $p \times q$-matrix C can be written in tensor form as

$$A_{jk} \otimes B_{\ell n} \otimes \delta_{k\ell} = A_{jk}B_{kn} = C_{jn} \quad , \tag{32}$$

which are $p \times q$ equations for the $p \times q$ components of C.

4.3 Symmetries of a tensor

When a tensor has two or more indices that have the same index range, one can compare the components belonging to these indices. If the sum or the difference of the components of exchanged indices are equal to zero, the tensor is called antisymmetric or symmetric in these indices. That means for a tensor $t_{jk\ell m}$:

if $t_{jk\ell m} = t_{kj\ell m} =: t_{(jk)\ell m} \quad \Rightarrow$ the tensor is called symmetric in j and k

if $t_{jk\ell m} = -t_{jkm\ell} =: t_{jk[\ell m]} \quad \Rightarrow$ the tensor is called antisymmetric in ℓ and m

A first example is the Kronecker delta. Since $\delta_{jk} = \delta_{kj}$, it is symmetric. As a second example I want to discuss the tensor a_{jk} of eq. (20). It has the components $a_{11} = a_{22} = 0.87$ and $a_{12} = -a_{21} = 0.48$. The last part, $a_{12} = -a_{21}$, hints, that it might be antisymmetric. But the relation for the tensor to be antisymmetric in one of its index pairs, i.e. $a_{jk} = -a_{kj}$, has to hold for **all** possible values of the indices. So it has to hold for the case $k = j$, too, which means that for a_{jk} to be antisymmetric, the diagonal element, i.e. the elements with

$k = j$ have to vanish[6]: $a_{\underline{jj}} = -a_{\underline{jj}} = 0$. But this is not the case in our second example: $a_{11} = a_{22} = 0.87 \neq 0$.

The third example will be the Riemann tensor, which describes the curvature of a manifold[7]. For the example we will assume, that the Riemann tensor describes a surface, that means its indices can have only two values: 1 or 2. Then the whole Riemann tensor can only have a single independent component, as it has the symmetry properties:

$$R_{jk\ell m} = -R_{kj\ell m} = -R_{jkm\ell} = R_{[\ell m]\,[jk]} = R_{1212} \ . \tag{33}$$

The Riemann tensor is antisymmetric in the first pair of indices and in the second pair of indices. Therefore neither the first pair nor the second pair can have the same index. But with only two choices for the indices, there is only a single index combination for the first and for the second pair. And since the pairs are symmetric, that means we can exchange the first with the second pair, only a single independent index combination is left, as we see in eq. (33).

4.4 Special tensors: the Levi-Civita symbol

Another special tensor is the Levi-Civita symbol. It does not have a fixed number of indices as it should be completely antisymmetric in the maximum number of possible indices, which requires that it has as many indices as there are dimensions, i.e. index choices. In two dimensions it can be written as $\epsilon_{jk} = -\epsilon_{kj}$ with $\epsilon_{12} = 1$. In three dimensions we have three indices:

$$\epsilon_{jk\ell} = \epsilon_{\ell jk} = \epsilon_{k\ell j} = -\epsilon_{kj\ell} = -\epsilon_{\ell kj} = -\epsilon_{j\ell k} \qquad \text{with} \quad \epsilon_{123} = 1 \ . \tag{34}$$

In four dimensions we have $\epsilon_{\mu\nu\rho\sigma}$, but I will not write the $4! = 24$ index combinations explicitly. The Levi-Civita symbol is used to define volume elements and determinants of matrices. In [4] the Levi-Civita symbol is called permutation symbol.

4.5 Restrictions on coordinate transformations

The two special tensors (or symbols), the Kronecker delta and the Levi-Civita symbol are defined in any coordinate system. Since they are tensors, their components have to be transformed under coordinate transformations, but these transformed components have to have the same values as in the original system. How is that possible?

Writing the coordinate transformation eq. (20) for the Kronecker delta we get

$$\delta'_{jk} = a_{j\ell} a_{km} \delta_{\ell m} = a_{j\ell} a_{k\ell} \ , \tag{35}$$

which tells us, that the coordinate transformation seen as a matrix should be orthogonal. Since rotations are described by orthogonal matrices this requirement should not surprise us.

[6] The underlining of an index indicates, that this index should not be summed over.

[7] I will neither explain what curvature means, nor what a manifold is. This example should just illustrate the symmetry properties of a tensor.

For the coordinate transformation eq. (20) acting on the Levi-Civita symbol we need to specify the dimensions of our space. We will first assume a two dimensional space and then a three dimensional. For 2D we get

$$\epsilon'_{jk} = a_{j\ell} a_{km} \epsilon_{\ell m} = a_{j1} a_{k2} - a_{j2} a_{k1} = \epsilon_{jk} \det[a] \ , \tag{36}$$

which tells us, that the determinant of the rotation matrix should be +1, which means that we should only use true rotations without reflections. For 3D we get

$$\epsilon'_{jk\ell} = a_{jm} a_{kn} a_{jo} \epsilon_{mno} = \epsilon_{jk\ell} \det[a] \ , \tag{37}$$

which gives the same conclusion as in the 2D example.

4.5.1 Remark on the shortcomings of this introduction

What was left out in the whole discussion was the concept of the dual space. It helps to understand better the difference between row and column vectors. It clarifies the different conventions about upper and lower indices. It gives the Kronecker delta and the contraction a more natural meaning. It also goes much deeper into the mathematical foundations: much deeper than is needed to do understand the calculations that are required by this course.

But when we use Cartesian coordinates the vector space and its dual can be identified. Therefore I decided to leave out the concept of the dual space and just concentrate on the methods that are needed for the simple calculations of the homework of this course.

5 Invariants

If we think more carefully about these restrictions on coordinate systems we will find them rather natural. When we compare different coordinate systems we do not want that the measured length changes when we use different coordinates. The distance between points is something that should not depend on our way of looking at them. In the same way the volume should stay the same. This viewpoint summarizes passive transformations. Our restrictions describe valid changes of coordinates systems, where we do not change the definition of our length measurement.

Things that do not change are called invariants. But being invariant is a broader concept than being invariant only under coordinate transformations. As an example we consider a pendulum like in Fig. 5. The length of the pendulum is by definition constant, as is its mass. When the pendulum swings its weight $m\vec{g}$ is usually also assumed to be constant. We can easily see, that this is only an assumption, when we consider a 1000 km long pendulum mounted on a satellite and swinging in earths gravitational field. Then the direction of the gravitational force will change depending on the position of the pendulum and $m\vec{g}$ will no longer be constant.

We introduce now a coordinate system and look at the quantities that we need to describe our pendulum. At each time we have two points, that describe the position of the pendulum: the point where it is mounted and the point where the weight is. From these two points we can define the vector $\vec{\ell}$, that points from the mounting point O to the weight

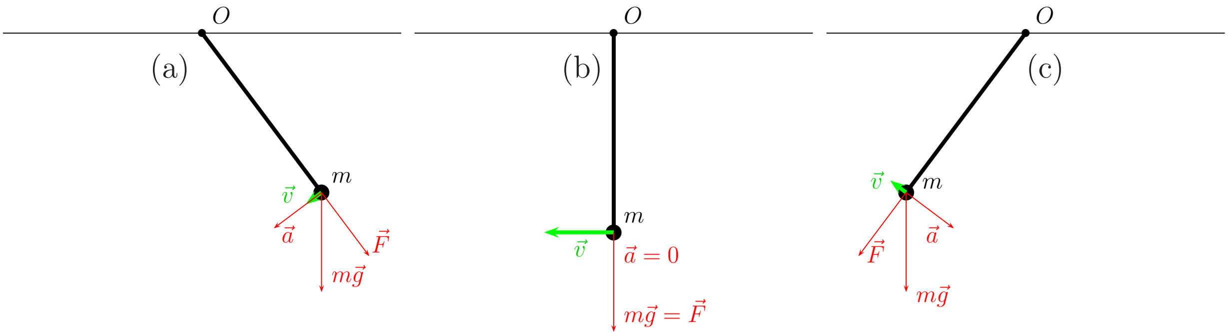

Figure 5: A rigid pendulum in three different positions. The forces are written in red, the velocity in green. The gravitational force is indicated by $m\vec{g}$, the acceleration by $\vec{a}$ and the centrifugal force by $\vec{F}$. In (a) the pendulum started very recently to move from its starting position. It has still nearly all of its potential energy and only a small velocity. The acceleration is parallel to the small velocity, so the velocity will increase. In (b) the velocity has increased and all potential energy has been converted to kinetic energy. When the pendulum swings on, as displayed in (c), its kinetic energy is transfered again to potential energy and the acceleration is anti parallel to the velocity, slowing the pendulum down.

m[8]. To understand the position of the pendulum we need additionally the direction of the gravitational attraction, which we call $\vec{g}$, the gravitational acceleration. Kinematically we also need the velocity $\vec{v}$ of the pendulum and its mass m. But with this information, we can completely describe it.

When we think now about the coordinate transformations discussed in the previous section, we can ask, which of the defining quantities of the pendulum stay the same under a coordinate transformation, that does not change our unit of length.

The directional quantities like $\vec{\ell}$ and $\vec{g}$ will obviously change when we describe the pendulum with a rotated coordinate system. But also the coordinates of the points O and m will change. So these quantities cannot be invariant. The fixed quantities of the pendulum, its mass m and its length have to stay invariant. The mass is described as a number and in the context of coordinates as a scalar quantity. The length[9] of the pendulum

$$\ell = |\vec{\ell}| = \sqrt{\vec{\ell}^{\,2}} = \sqrt{\vec{\ell}\cdot\vec{\ell}} = \sqrt{\ell_i \ell_i} \tag{38}$$

is the length of the vector $\vec{\ell} = \ell_i$. And here we see a connection between the invariants of the coordinate transformations and invariants in a physical system: since physics should be independent of our choice of a coordinate system only invariants under coordinate transformations can be invariants of a physical system.

[8]In this example I will use the letter m to describe the mass of the pendulum and to name the position of the mass. There is no confusion about mixing a mass with a position, but an advantage to use the same letter for the same object, since I do not need to write more letters to the pictures in Fig. 5, which would reduce its readability.

[9]The definition of the square of the length of a vector $|\vec{\ell}|^2 = \vec{\ell}\cdot\vec{\ell}$ needs the definition of a scalar product, which in turn requires the introduction of a metric. In the spirit of section 4.5.1 I will assume, that we know, how to calculate the length of a vector and use this understanding as an implicit definition of the scalar product and the metric.

Any quantity that does not change under coordinate transformations is called a scalar. We can think of the scalar as a tensor without an index: when we change our coordinates and transform the tensor according to eq. (25) we have no factor $\partial x'_j/\partial x_k$ for a scalar, since we have no index j that we have to transform.

We can construct many scalar quantities, but not all of them will be invariants for the physical system, not all of them will stay constant over time. One example for such a scalar quantity is the product of the vector $\vec{\ell}$ with the gravitational force $m\vec{g}$:

$$E_{\text{pot}} := -m\,\vec{g}\cdot\vec{\ell} \tag{39}$$

describes the potential energy of the pendulum. This quantity will be the same in all equivalent coordinate systems[10]. But the potential energy is not the same for the states of the pendulum in part (a) and part (b) of Fig. 5.

Another scalar quantity is the kinetic energy

$$E_{\text{kin}} := \tfrac{1}{2}\,m\,\vec{v}^2 \quad . \tag{40}$$

The vector $\vec{v}$ changes under coordinate transformations, but its scalar length $|\vec{v}|$ does not. Over time E_{kin} changes as can be seen in part (a) and part (b) of Fig. 5. Now the sum of both quantities

$$E_{\text{tot}} := E_{\text{pot}} + E_{\text{kin}} \quad . \tag{41}$$

is obviously also a scalar quantity an will not change under coordinate transformations. But this time, E_{tot} is also a physical invariant, a quantity that does not change over time.

6 Galilean transformations

The coordinate transformations we discussed in sec. 3.1 and 3.2 are part of the Galilean transformations. Additionally one has transformations including an universal time. The transformation of time analogous to eq. (6) is the time shift

$$t' = t + \Delta t \quad , \tag{42}$$

where I wrote the formal parameter Δt instead of the explicit example values $+2$ and -1 in eq. (6).

In our three spacial dimensions the translations can be written formally as

$$x' = x + \Delta x \quad , \quad y' = y + \Delta y \quad \text{and} \quad z' = z + \Delta z \quad . \tag{43}$$

These three equations can be written shorter in vector form

$$\vec{x}' = \vec{x} + \Delta\vec{x} \tag{44}$$

or index form

$$x'_j = x_j + \Delta x_j \quad . \tag{45}$$

[10] We will call all coordinate systems that do not change the length measurements equivalent.

original	time shift	space shift	rotation	boost
(t, x, y, z)	$(t + \Delta t, x, y, z)$	$(t, x + \Delta x, y, z)$	(t, x_R, y_R, z_R)	$(t, x + w_x t, y, z)$
$(t, \vec{x})$	$(t + \Delta t, \vec{x})$	$(t, \vec{x} + \Delta\vec{x})$	$(t, \mathbf{R} \cdot \vec{x})$	$(t, \vec{x} + \vec{w}\, t)$
(t, x_j)	$(t + \Delta t, x_j)$	$(t, x_j + \Delta x_j)$	$(t, \mathbf{R}_{jk} x_k)$	$(t, x_j + w_j t)$

Table 1: The Galilean transformation in the three ways of writing them. The explicit form is listed in the first line, where I only write the space shift and the boost for the $\hat{x}$-direction, because it would not fit the page width. The vector form is given in the second line and the index form in the last line. For the rotation in the last line, you should remember Einsteins summation convention given in eq. (20) and described after the equation.

The rotations of sec. 3.2 are written in eqs. (21a), (21b), and (21c), where the parameters $b_1 = b_2 = b_3 = 0$ have to be set to zero. The index form of the rotations can be found in eq. (22), again setting $b_j = 0$, and the matrix form of the rotations can be found in eq. (18).

The last Galilean transformation is motivated by the observation, that in every day life circumstances we can just add velocities like normal vectors. The names of these transformations are boosts. You can remember the name by imagining, that the transformation is achieved by giving one coordinate system a boost, that is a push, so that it glides with constant velocity into the direction in which you pushed it. Formally one can write these three equations[11] as

$$x' = x + w_x t \quad , \quad y' = y + w_y t \quad \text{and} \quad z' = z + w_z t \ , \tag{46}$$

or easier in vector form

$$\vec{x}' = \vec{x} + \vec{w}\, t \tag{47}$$

or index form

$$x'_j = x_j + w_j t \ . \tag{48}$$

You can imagine the boosts also as a normal translation that depends linearly on time: at time t_1 you make the translation with $\Delta\vec{x} = \vec{w}\, t_1$ and at time t_2 you make the translation with $\Delta\vec{x} = \vec{w}\, t_2$. In that sense the boost is not different from the normal translations.

6.1 The pendulum under Galilean transformations

Looking again at the pendulum of Fig. 5 we can apply all three types of Galilean transformations to the quantities defining the pendulum: to the positions of O and m, to the vectors $\vec{\ell}$, $\vec{v}$, $\vec{g}$, $\vec{F}$, and $\vec{a}$, and the scalar quantities m and ℓ.

[11] I use the letter w for the velocity of the boost, as I already used the usual letter v for the velocity of the pendulum, that I want to discuss later again.

	untransformed	time shift	space shift	rotation	boost
$\vec{O}$	$\vec{O}_0 = (O_x, O_y, O_z)$	$\vec{O}_T = \vec{O}_0$	$\vec{O}_S = \vec{O}_0 + \Delta\vec{x}$	$\vec{O}_R = \mathbf{R}\cdot\vec{O}_0$	$\vec{O}_B = \vec{O}_0 + \vec{w}\,t$
$\vec{m}$	$\vec{m}_0 = (m_x, m_y, m_z)$	$\vec{m}_0(t+\delta t)$	$\vec{m}_S = \vec{m}_0 + \Delta\vec{x}$	$\vec{m}_R = \mathbf{R}\cdot\vec{m}_0$	$\vec{m}_B = \vec{m}_0 + \vec{w}\,t$
$\vec{\ell}$	$\vec{\ell}_0 = \vec{m}_0 - \vec{O}_0$	$\vec{\ell}_0(t+\delta t)$	$\vec{\ell}_S = \vec{\ell}_0$	$\vec{\ell}_R = \mathbf{R}\cdot\vec{\ell}_0$	$\vec{\ell}_B = \vec{\ell}_0$
$\vec{v}$	$\vec{v}_0 = (v_x, v_y, v_z)$	$\vec{v}_0(t+\delta t)$	$\vec{v}_S = \vec{v}_0$	$\vec{v}_R = \mathbf{R}\cdot\vec{v}_0$	$\vec{v}_B = \vec{v}_0 + \vec{w}$
$\vec{g}$	$\vec{g}_0 = (g_x, g_y, g_z)$	$\vec{g}_T = \vec{g}_0$	$\vec{g}_S = \vec{g}_0$	$\vec{g}_R = \mathbf{R}\cdot\vec{g}_0$	$\vec{g}_B = \vec{g}_0$
$\vec{F}$	$\vec{F}_0 = \frac{(m\,\vec{g}_0\cdot\vec{\ell}_0)}{(\vec{\ell}_0\cdot\vec{\ell}_0)}\vec{\ell}_0$	$\vec{F}_0(t+\delta t)$	$\vec{F}_S = \vec{F}_0$	$\vec{F}_R = \mathbf{R}\cdot\vec{F}_0$	$\vec{F}_B = \vec{F}_0$
$\vec{a}$	$\vec{a}_0 = m\,\vec{g}_0 - \vec{F}_0$	$\vec{a}_0(t+\delta t)$	$\vec{a}_S = \vec{a}_0$	$\vec{a}_R = \mathbf{R}\cdot\vec{a}_0$	$\vec{a}_B = \vec{a}_0$
ℓ	$\ell_0 = \sqrt{\vec{\ell}_0\cdot\vec{\ell}_0}$	$\ell_T = \ell_0$	$\ell_S = \ell_0$	$\ell_R = \ell_0$	$\ell_B = \ell_0$
m	$m_0 = m$	$m_T = m_0$	$m_S = m_0$	$m_R = m_0$	$m_B = m_0$
E_{pot}	$E_{\text{p},0} = -m\,\vec{g}_0\cdot\vec{\ell}_0$	$E_{\text{p},0}(t+\delta t)$	$E_{\text{p},S} = E_{\text{p},0}$	$E_{\text{p},R} = E_{\text{p},0}$	$E_{\text{p},B} = E_{\text{p},0}$
E_{kin}	$E_{\text{k},0} = \frac{1}{2}\,m\,\vec{v}_0^2$	$E_{\text{k},0}(t+\delta t)$	$E_{\text{k},S} = E_{\text{k},0}$	$E_{\text{k},R} = E_{\text{k},0}$	$E_{\text{k},B} \neq E_{\text{k},0}$
E_{tot}	$E_{\text{t},0} = E_{\text{pot}} + E_{\text{kin}}$	$E_{\text{t},T} = E_{\text{t},0}$	$E_{\text{t},S} = E_{\text{t},0}$	$E_{\text{t},R} = E_{\text{t},0}$	$E_{\text{t},B} \neq E_{\text{t},0}$

Table 2: The Galilean transformation of the quantities that describe the pendulum of Fig. 5. The first block includes the coordinates of the pendulum $\vec{m}$, which changes with time, and the mount of the pendulum $\vec{O}$, which does not change with time. The second block includes the vector of the pendulum $\vec{\ell}$, which is the difference between the position of the weight and the mount, the velocity of the pendulum $\vec{v}$, the gravitational acceleration $\vec{g}$, the centrifugal force $\vec{F}$, and the acceleration of the weight $\vec{a}$. Only the gravitational acceleration $\vec{g}$ is constant in time. The third block contains the scalar quantities of the pendulum: the length ℓ and the mass m. These quantities change neither with a coordinate transformation nor with time. The fourth block includes the energies of the pendulum.

This time I do not give numbers as an example. I leave the transformations and the describing quantities in parametric form. But I unify the notation for the translations in time and space and use an index to distinguish them. I use the subscript $_T$ for a time translated quantity, the subscript $_S$ for a spacial translated quantity, the subscript $_R$ for a rotated quantity, and the subscript $_B$ for a boosted quantity. The transformations are summarized in table 1.

The first block in table 2 describes the coordinates of the mounting point $\vec{O}$ and the position of the weight $\vec{m}$. These vectorial quantities transform exactly like in table 1. Since the pendulum swings, the position of the weight changes with time. When we look at the pendulum at a different time, the position will also be different, which is indicated by writing the time shifted $\vec{m}_T$ as a function of the time t and of the time shift Δt.

The second block contains vectors that are not position vectors. The vector describing the arm of the pendulum, $\vec{\ell} = \vec{m} - \vec{O}$ is a difference of position vectors. Since the position vector $\vec{m}$ changes with time, but the position vector $\vec{O}$ does not, $\vec{\ell}$ inherits the same time dependence as $\vec{m}$, which is indicated by writing $\vec{\ell}_T$ as a function of the time t and of the time shift Δt. Any spacial shift leaves $\vec{\ell}$ invariant, as one can see from table 2. Remember that the boost is also a spacial shift, but a spacial shift that is different for different times. Only the rotation changes $\vec{\ell}$ by rotating it like any other vector:

$$\vec{\ell}_R = \vec{m}_R - \vec{O}_R = \mathbf{R} \cdot \vec{m}_0 - \mathbf{R} \cdot \vec{O}_0 = \mathbf{R} \cdot (\vec{m}_0 - \vec{O}_0) = \mathbf{R} \cdot \vec{\ell}_0 \tag{49}$$

or

$$\ell_{Rj} = m_{Rj} - O_{Rj} = \mathbf{R}_{jk} m_{0k} - \mathbf{R}_{jk} O_{0k} = \mathbf{R}_{jk}(m_{0k} - O_{0k}) = \mathbf{R}_{jk}\ell_{0k} \ . \tag{50}$$

The velocity of the weight $\vec{v}$ is also a difference of two position vectors. But now the difference is taken at different times[12]:

$$\vec{v}(t) = \lim_{\delta t \to 0} \frac{\vec{m}(t + \delta t) - \vec{m}(t)}{\delta t} \qquad \text{or} \qquad v_j(t) = \lim_{\delta t \to 0} \frac{m_j(t + \delta t) - m_j(t)}{\delta t} \ . \tag{51}$$

Therefore the boost, which changes the coordinates gives a different $\vec{v}_B$:

$$\begin{aligned} \vec{v}_B(t) &= \lim_{\delta t \to 0} \frac{[\vec{m}_0(t + \delta t) + \vec{w} \cdot (t + \delta t)] - [\vec{m}_0(t) + \vec{w} \cdot (t)]}{\delta t} \\ &= \lim_{\delta t \to 0} \frac{\vec{m}_0(t + \delta t) - \vec{m}_0(t)}{\delta t} + \vec{w}\, \frac{(t + \delta t) - (t)}{\delta t} = \vec{v}_0(t) + \vec{w} \ . \end{aligned} \tag{52}$$

The rotation of the velocity $\vec{v}_R$ is obtained in the same way as the rotation of the arm of the pendulum $\vec{\ell}_R$ described in eq. (49) and eq. (50).

The acceleration $\vec{g}$ is in principle defined as the differential of the velocity. Since the effect of the gravitational acceleration of the pendulum could only be seen directly if the pendulum was not mounted, we can only argue, that $\vec{g}$ should change like any second derivative with respect to time. Or we can argue, that $\vec{g}$ should not change, like the coordinate axes themselves, as we are keeping the gravitational acceleration fixed for the description of the pendulum. For the change of a second derivative with a boost we discuss the change of $\vec{a}$ under a boost. The definition of $\vec{a}$ in terms of the velocity $\vec{v}$ uses the same formalism as the definition of the $\vec{v}$ velocity in terms of the position $\vec{m}$:

$$\vec{a}(t) = \lim_{\delta t \to 0} \frac{\vec{v}(t + \delta t) - \vec{v}(t)}{\delta t} \qquad \text{or} \qquad a_j(t) = \lim_{\delta t \to 0} \frac{v_j(t + \delta t) - v_j(t)}{\delta t} \ . \tag{53}$$

Calculating the boosted acceleration

$$\vec{a}_B(t) = \lim_{\delta t \to 0} \frac{\vec{v}_B(t + \delta t) - \vec{v}_B(t)}{\delta t} = \lim_{\delta t \to 0} \frac{[\vec{v}_0(t + \delta t) + \vec{w}] - [\vec{v}_0(t) + \vec{w}]}{\delta t} = \vec{a}_0(t) \tag{54}$$

works in a similar way as the boosted velocity eq. (52), with the difference, that the boost does not change the acceleration. So it does not matter, if we treat the gravitational acceleration $\vec{g}$ as an acceleration or a coordinate axis.

[12]The limit for infinitesimal small differences is called differentiation.

The centrifugal force $\vec{F}$ is calculated from vectors that do not change under spacial translations. Therefore it will not change under spacial translations, either. Under temporal translations $\vec{F}$ will change in a similar way like all the quantities that depend on the position of the pendulum. That $\vec{F}$ changes under rotations in the same way as $\vec{\ell}$ follows from the invariance of the scalar product, which is defined by the Kronecker delta. The invariance of the Kronecker delta under rotations was discussed in eq. (35).

The third block of table 2 describes the defining scalar quantities of the pendulum. The length of the arm of the pendulum cannot change due to the physical requirements of the pendulum: if ℓ would change in time, the pendulum would not be rigid. And if m could change we would also not have the pendulum we discussed. The constancy of ℓ gives a restriction, how the weight can move and makes this pendulum a one (two) dimensional problem, although it moves in two (three) spacial dimensions. Assuming the rigidity, i.e. that the length ℓ does not change, one can replace the vector $\vec{m}$ by the angles describing its direction.

The last block of table 2 describes the energies of the pendulum. The potential energy depends on the direction of the gravitational acceleration $\vec{g}$ and the arm of the pendulum $\vec{\ell}$. The gravitational acceleration is constant by assumption, but the arm of the pendulum swings and so the potential energy changes with time. The potential energy does not change under a spacial shift since the direction of the arm $\vec{\ell}$ is the difference of two position vectors and the gravitational acceleration is constant by assumption. Rotations leave a scalar product of two vectors invariant, so they do not change the potential energy. Since boost do not change the vectors defining the potential energy, they do not change the potential energy, either.

The kinetic energy is given by the length of the velocity vector $\vec{v}$ and changes periodically with time. But spacial translations and rotations do not change the length of a vector and hence do not change the kinetic energy. But the boost adds its velocity to the velocity of the pendulum and so it changes the kinetic energy.

The total energy is the sum of potential and kinetic energy. Since both, potential and kinetic energy are not affected by spacial translations and by rotations, these transformations will not affect the total energy either. The boosts do not change the potential energy, but change the kinetic energy and so the total energy will not be constant under boosts. Both the potential and the kinetic energy change with time, but they change in a very special way: the kinetic energy increases by the same amount that the potential energy decreases and vice versa. This makes the total energy a constant of motion for the pendulum.

This splitting of the total energy in potential and kinetic part is the foundation for analytical mechanics. The advantage of analytical mechanics is the usage of scalar quantities, the potential and the kinetic energy, instead of using forces and other vector like quantities, that transform under translations and rotations.

7 Four dimensional spacetime

The difference between Galilean transformations and Lorentz transformations lies only in the boosts. Whereas Galilean transformations use the notion of an universal time, Lorentz transformations recognize the connection between time and space, the relativity of time measurements. Therefore the first step into Lorentz transformations is the introduction of the four-vector notation, which unifies time and space formally.

7.1 Four-vectors

In section 2 we introduced the vector by giving the Cartesian coordinates in space. For the four-vector we just have to include the time that is measured together with the position and then we have the typical four-vector. But for time we use different physical units compared to the length measurements in space. The simplest solution is to take the time coordinate as the distance light travels in the time that we want to write as the time coordinate:

$$x^0 = ct \ . \tag{55}$$

Then there is no problem connecting time and space into the same mathematical object:

$$x^\mu = (ct, \vec{x}) = (x^0, x^1, x^2, x^3) \qquad \text{or} \qquad x^\mu = \begin{pmatrix} ct \\ \vec{x} \end{pmatrix} = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \ . \tag{56}$$

As we can see it does not make a difference, whether we write the four-vector as a row vector or as a column vector. And like in the three dimensional case it will be convenient to write the four-vector in index notation, discussed in section 4.

In index notation there is now a difference to the three dimensional case. For a four-vector there exists a distinction between upper and lower indices. In special relativity this distinction is just a convention, but in Riemannian geometry and general relativity this index convention is really necessary. In special relativity it also helps to prevent mistakes. An upper index is called contravariant and a lower index is called covariant. Since a vector has only a single index, it has to be either contravariant or covariant. But a general tensor with more than a single index can have contravariant and covariant indices.

Einsteins summation convention is slightly different than in three dimensions: each index that appears once as an upper index and once as a lower index in one term should be summed over. And this is the part, where the convention about contravariant and covariant indices can prevent mistakes. If you encounter in your calculation a term that has two similar indices and both are upper or lower indices, you know, that you either named a different index with the same name, or that you tried to multiply two incompatible quantities.

In section 5 we we found that most invariant quantities are scalars. In the context of special relativity this is true, too. And a scalar is easy to see: it does not have an index. That can come from the contraction of two indices, which are summed over by Einsteins summation convention. So one of the indices has to be a contravariant index and the other a covariant index. But how can one then measure the length of a three dimensional vector, eq. (56), at a certain time?

7.2 The metric

The metric is the tool that allows the length measurement. In general relativity the metric becomes the object, that describes the local structure of space and time. In special relativity this is the case, too, but the metric in Cartesian coordinates is constant and does not depend on the place and on the time where the measurement is made.

In Cartesian coordinates we know, that the length measurement, eq. (38), means we have to sum the squares of the spacial components of the vector to get the square of the length of the vector we want to measure. So the spacial components of the metric in Cartesian coordinates have to be the Kronecker delta. The rest of the metric is again convention. In astronomy people want to measure positive distances and take the metric to have positive space components. In particle physics we like to have positive energies and therefore we take the time component to be positive[13]:

$$g_{\mu\nu} \;:=\; \{\, g_{00} = +1\,; g_{0j} = 0\,; g_{jk} = -\delta_{jk} \text{ with } \mu,\nu = 0\ldots 3 \text{ and } j,k = 1\ldots 3\,\} \;. \tag{57}$$

This metric allows now the connection between contravariant and covariant indices.

We can lower indices by contracting a contravariant index with the metric

$$a_\mu = g_{\mu\nu} a^\nu \;=\; \{\, +a^\mu \text{ for } \mu = 0 \text{ and } -a^\mu \text{ for } \mu = 1\ldots 3\,\} \;. \tag{58}$$

We can also lower an index from a multi-index tensor:

$$\begin{aligned} t_\mu{}^{\rho\sigma} = g_{\mu\nu} t^{\nu\rho\sigma} \qquad &\text{or} \qquad t_\mu{}^\rho{}_\sigma = g_{\mu\nu} t^{\nu\rho}{}_\sigma = g_{\sigma\nu} t_\mu{}^{\rho\nu} \\ \text{or} \qquad t_{\mu\rho\sigma} &= g_{\mu\nu} t^\nu{}_{\rho\sigma} = g_{\rho\nu} t_\mu{}^\nu{}_\sigma = g_{\sigma\nu} t_{\mu\rho}{}^\nu \;, \end{aligned} \tag{59}$$

but here we have to keep the index at its position. Note that the tensor t can describe the same physical object, but its realisations with different index positions are *not the same*:

$$t^{\mu\rho\sigma} \neq t_\mu{}^{\rho\sigma} \neq t^\mu{}_\rho{}^\sigma \neq t^{\mu\rho}{}_\sigma \neq t_{\mu\rho}{}^\sigma \neq t_\mu{}^\rho{}_\sigma \neq t^\mu{}_{\rho\sigma} \neq t_{\mu\rho\sigma} \;. \tag{60}$$

For raising indices we need the inverse metric

$$g^{\mu\nu} \;:=\; \{\, g^{00} = +1\,; g^{0j} = 0\,; g^{jk} = -\delta_{jk} \text{ with } \mu,\nu = 0\ldots 3 \text{ and } j,k = 1\ldots 3\,\} \;, \tag{61}$$

which gives the four dimensional Kronecker delta when contracted with the regular metric:

$$g_{\mu\nu} g^{\nu\rho} = \delta_\mu^\rho \;. \tag{62}$$

The index position in the four dimensional Kronecker delta is not important, since it becomes the metric, when one index is lowered or the inverse metric when one index is raised. And both, the metric and the inverse metric, are symmetric, so it does not matter which of the indices is in front of the other. In the same way it does not matter which index of the (inverse) metric we use to lower (raise) an index, since the (inverse) metric is symmetric in its two indices.

[13] In quantum mechanics position and momentum are conjugate to each other. In a similar same way we understand time as being conjugate to energy.

7.3 Length measurement

In section 5 we discussed the definition of length in Cartesian coordinates, eq. (38). We have to generalize this result to our four dimensional spacetime. Looking at the pendulum, Fig. 5, we see, that the length is defined by the difference between the points O and m, which we can also write with their four dimensional spacetime vectors (in row vector form)

$$O^\mu = (ct, 0, 0, 0) \qquad \text{and} \qquad m^\mu = (ct, m_x, m_y, m_z) \ . \tag{63}$$

This means we have the components of O

$$O^0 = ct \qquad \text{and} \qquad O^1 = O^2 = O^3 = 0 \tag{64}$$

and the components of m

$$m^0 = ct \qquad \text{and} \qquad m^1 = m_x \ , \quad m^2 = m_y \ , \quad \text{and} \quad m^3 = m_z \ . \tag{65}$$

We have to take the same time t for both points, as we want to measure the length, and the length measurement is **defined** to be at equal times. The four-vector difference between O and m is

$$\overrightarrow{Om}^\mu = (ct, m_x, m_y, m_z) - (ct, 0, 0, 0) = (0, m_x, m_y, m_z) = (0, \vec{\ell}) =: d_s^\mu \ , \tag{66}$$

where we identified the three dimensional coordinates with the space-like four dimensional coordinates: $\vec{\ell} = (m_x, m_y, m_z)$. In the three dimensional case of eq. (38) we also had a metric, which was just the Kronecker delta in Cartesian coordinates. And the square of the length was written as

$$\ell^2 = \ell_i \ell_i = \ell_j \ell_k \delta_{jk} \ . \tag{67}$$

A similar calculation with four-vectors gives

$$g_{\mu\nu} d_s^\mu d_s^\nu = (d_s^0)^2 - (d_s^1)^2 - (d_s^2)^2 - (d_s^3)^2 = -\delta_{jk} \ell_j \ell_k = -\ell^2 \ , \tag{68}$$

which gives a negative number for the square of a length. That means, we have to **define** the length measurement in four dimensions as

$$\ell := \sqrt{-\left[g_{\mu\nu} d_s^\mu d_s^\nu\right]_{\text{space like part}}} \ . \tag{69}$$

7.4 Time measurement

The definition of a time measurement seems to be a trivial task: We just take a clock and measure the time that passes. And nothing is wrong with this. But when we look at the time measurement having the principle of relativity in mind, we have to ask, how can we measure the time between two events? And now the answer is: in order to measure the time uniquely between two events, both events have to be at the same location. Then we can measure the time that passes in that location and that will give the unique answer.

We can also use the four-vector notation for the two events A and B:

$$A^\mu = (a_t = ct_A, a_x, a_y, a_z) \qquad \text{and} \qquad B^\mu = (b_t = ct_B, b_x, b_y, b_z) \ . \tag{70}$$

In order to measure the time between A and B they have both to be at the same location, which means

$$a_x = b_x \quad , \qquad a_y = b_y \quad , \quad \text{and} \qquad a_z = b_z \ , \tag{71}$$

which simplifies the difference

$$\overrightarrow{AB}^\mu = (b_t, b_x, b_y, b_z) - (a_t, a_x, a_y, a_z) = (c(t_B - t_A), 0, 0, 0) = (c\Delta t, 0, 0, 0) =: d_t^\mu \ . \tag{72}$$

Using the metric as with the length measurement we can **define** the time measurement as

$$c\Delta t := \sqrt{+\,[g_{\mu\nu} d_t^\mu d_t^\nu]_{\text{time like part}}} \ . \tag{73}$$

This definition unifies space and time and motivates the definition of the four-distance squared d^2 between A and B:

$$d^\mu := B^\mu - A^\mu \qquad \text{and} \qquad d^2 := g_{\mu\nu} d^\mu d^\nu \ . \tag{74}$$

7.5 Lorentz transformations

The Galilean transformations listed in Table 1 can be understood as the linear transformations that leave the length measurement, eq. (38), invariant. Since time is not changed by the Galilean transformations, also the time measurement is invariant under Galilean transformations.

In this context Lorentz transformations can be understood as the linear transformations that leave the four-distance squared, eq. (74), invariant. This feature already determines the Lorentz transformations. This approach to the Lorentz transformations is used in [1] and in the lecture.

Using four-vectors and index notation we can write the Lorentz transformation $\Lambda(v)$ as the transformation from the frame O with Cartesian coordinates x^μ to the frame O' with Cartesian coordinates x'^μ which moves with the constant velocity $\vec{v} = (v, 0, 0)$ in $\hat{x}$-direction with respect to the frame O:

$$x'^\mu = \Lambda^\mu{}_\nu \, x^\nu \ . \tag{75}$$

This can be written out explicitly into four equations

$$\begin{aligned}
x'^0 &= \Lambda^0{}_0\, x^0 + \Lambda^0{}_1\, x^1 + \Lambda^0{}_2\, x^2 + \Lambda^0{}_3\, x^3 \\
x'^1 &= \Lambda^1{}_0\, x^0 + \Lambda^1{}_1\, x^1 + \Lambda^1{}_2\, x^2 + \Lambda^1{}_3\, x^3 \\
x'^2 &= \Lambda^2{}_0\, x^0 + \Lambda^2{}_1\, x^1 + \Lambda^2{}_2\, x^2 + \Lambda^2{}_3\, x^3 \\
x'^3 &= \Lambda^3{}_0\, x^0 + \Lambda^3{}_1\, x^1 + \Lambda^3{}_2\, x^2 + \Lambda^3{}_3\, x^3 \ .
\end{aligned} \tag{76}$$

The Lorentz transformation affects only the time coordinate and the coordinates in the direction of the velocity. Since we choose the velocity to point into the $\hat{x}$-direction the coordinates in $\hat{y}$-direction and $\hat{z}$-direction are not changed:

$$\begin{gathered}
x'^2 = x^2 \qquad \text{and} \qquad x'^3 = x^3 \\
< \text{or} \qquad \Lambda^2{}_2 = \Lambda^3{}_3 = 1 \qquad \text{and} \\
\Lambda^0{}_2 = \Lambda^0{}_3 = \Lambda^1{}_2 = \Lambda^1{}_3 = \Lambda^2{}_0 = \Lambda^2{}_1 = \Lambda^2{}_3 = \Lambda^3{}_0 = \Lambda^3{}_1 = \Lambda^3{}_2 = 0 \quad .
\end{gathered} \tag{77}$$

Requiring the invariance of eq. (74) for x'^μ and x^μ we get $(x')^2 = (x)^2$. Writing x'^μ explicitly by using eq. (76) with the restrictions eq. (77) we get

$$\begin{aligned}
& (x'^0)^2 - (x'^1)^2 - (x'^2)^2 - (x'^3)^2 \\
= \; & (\Lambda^0{}_0\, x^0 + \Lambda^0{}_1\, x^1)^2 - (\Lambda^1{}_0\, x^0 + \Lambda^1{}_1\, x^1)^2 - (x^2)^2 - (x^3)^2 \\
= \; & \quad (\Lambda^0{}_0)^2 (x^0)^2 + 2\Lambda^0{}_0 \Lambda^0{}_1 x^0 x^1 + (\Lambda^0{}_1)^2 (x^1)^2 \\
& -(\Lambda^1{}_0)^2 (x^0)^2 - 2\Lambda^1{}_0 \Lambda^1{}_1 x^0 x^1 - (\Lambda^1{}_1)^2 (x^1)^2 \;\; - (x^2)^2 - (x^3)^2 \\
= \; & \quad [(\Lambda^0{}_0)^2 - (\Lambda^1{}_0)^2](x^0)^2 + 2[\Lambda^0{}_0 \Lambda^0{}_1 - \Lambda^1{}_0 \Lambda^1{}_1] x^0 x^1 \\
& +[(\Lambda^0{}_1)^2 - (\Lambda^1{}_1)^2](x^1)^2 \;\; - (x^2)^2 - (x^3)^2 \\
\doteq \; & (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \ .
\end{aligned} \tag{78}$$

Since there is no term proportional $x^0 x^1$ on the right hand side of $\doteq$ we know that its coefficient has to vanish. This gives the equation

$$\Lambda^0{}_0 \Lambda^0{}_1 - \Lambda^1{}_0 \Lambda^1{}_1 = 0 \ . \tag{79}$$

The coefficients for $(x^0)^2$ and $(x^1)^2$ have to be the same in eq. (78). So we get two more equations:

$$(\Lambda^0{}_0)^2 - (\Lambda^1{}_0)^2 = 1 \tag{80}$$
$$\text{and} \qquad (\Lambda^0{}_1)^2 - (\Lambda^1{}_1)^2 = -1 \ . \tag{81}$$

Remembering the meaning of $x^0 = ct$ being the time we can assume that the transformation eq.(76) has to have $\Lambda^0{}_0 \neq 0$, so that we can solve eq.(79) to get

$$\Lambda^0{}_1 = \frac{\Lambda^1{}_0 \Lambda^1{}_1}{\Lambda^0{}_0} \ , \tag{82}$$

which we can put into eq.(81)

$$\left(\frac{\Lambda^1{}_0 \Lambda^1{}_1}{\Lambda^0{}_0}\right)^2 - (\Lambda^1{}_1)^2 = \left(\frac{\Lambda^1{}_1}{\Lambda^0{}_0}\right)^2 \left((\Lambda^1{}_0)^2 - (\Lambda^0{}_0)^2\right) = -1 \ . \tag{83}$$

Using eq.(80) we get

$$\left(\frac{\Lambda^1{}_1}{\Lambda^0{}_0}\right)^2 = 1 \quad \text{or} \qquad \Lambda^1{}_1 = \pm \Lambda^0{}_0 \ . \tag{84}$$

We can now stop and think of the meaning of the signs of $\Lambda^0{}_0$ and $\Lambda^1{}_1$. When $\Lambda^0{}_0 > 0$ time in both frames, O and O', are measured in the same direction, which is what we experience in everyday life: both physicists, the one at rest and the one moving get older and not younger. If $\Lambda^0{}_0 < 0$ one of them would have to get younger when the other one gets older. The sign of $\Lambda^1{}_1$ is a convention of the coordinate systems. If $\Lambda^1{}_1 < 0$ the physicist in O' measures the coordinate in $\hat{x}$-direction into the opposite direction compared to the physicist in O. This is not convenient when we want to compare systems which are moving only very slowly with respect to each other. So the usual convention is to choose ${\Lambda^1}_1 > 0$, too. This determines now

$$ {\Lambda^1}_1 = {\Lambda^0}_0 \tag{85} $$

and therefore from eq. (79) or eq. (82) also

$$ {\Lambda^0}_1 = {\Lambda^1}_0 \ . \tag{86} $$

This leaves only one equation, eq. (80) or eq. (81). They can be solved by parameterizing either both elements, ${\Lambda^0}_0$ and ${\Lambda^1}_0$, by two functions and one parameter η, called rapidity,

$$ {\Lambda^0}_0 = \cosh\eta \qquad \text{and} \qquad {\Lambda^1}_0 = \pm\sinh\eta \ , \tag{87} $$

or by parameterizing the ratio by β,

$$ \frac{{\Lambda^1}_0}{{\Lambda^0}_0} = \pm\beta \ , \tag{88} $$

to get

$$ ({\Lambda^0}_0)^2 - (\pm\beta{\Lambda^0}_0)^2 = ({\Lambda^0}_0)^2(1-\beta^2) = 1 \tag{89} $$

and calculating then both elements

$$ {\Lambda^0}_0 = \frac{1}{\sqrt{1-\beta^2}} =: \gamma \qquad \text{and} \qquad {\Lambda^1}_0 = \pm\gamma\beta \ . \tag{90} $$

The sign in eq.(87) or eq.(88) tells the direction of the motion between the frames O and O': when the Lorentz transformation from O to O' has the minus sign, then the transformation back from O' to O has the plus sign and vice versa.

The meaning of the parameter β is visible, when we look at the motion of the origin of frame O' in our frame O. When the origin of O' moves with the speed v in positive $\hat{x}$-direction, we see it at the time t_0 at the position x_0 and at time $t_1 = t_0 + \Delta t$ at the position $x_1 = x_0 + v\Delta t$. But the coordinates in O' stay the same at both times, $x'_1 = x'_0$, since it is still the origin of O'. This gives us the equation

$$ {\Lambda^1}_0\, ct_0 + {\Lambda^1}_1\, x_0 = {\Lambda^1}_0\, ct_1 + {\Lambda^1}_1\, x_1 = {\Lambda^1}_0\, c(t_0+\Delta t) + {\Lambda^1}_1\,(x_0 + v\Delta t) \tag{91} $$

or

$$ 0 = ({\Lambda^1}_0\, c + {\Lambda^1}_1\, v)\Delta t = \gamma(\pm\beta c + v)\Delta t \ . \tag{92} $$

Here we see the usual choice for the Lorentz transformation from O to O': one chooses the minus sign and gets the transformation parameter

$$ \beta = \frac{v}{c} \ . \tag{93} $$

This gives finally the form of the Lorentz transformations in $\hat{x}$-direction that most people are familiar:

$$ \begin{array}{rcl} ct' &=& \gamma\,(ct - \beta x) \\ x' &=& \gamma\,(x - \beta ct) \\ y' &=& y \\ z' &=& z \end{array} \qquad \text{or} \qquad \begin{array}{rcl} t' &=& \gamma\left(t - \frac{v\cdot x}{c^2}\right) \\ x' &=& \gamma\,(x - v\cdot t) \\ y' &=& y \\ z' &=& z \ . \end{array} \tag{94} $$

7.6 Four-momentum

Up to now we talked only about general frames. The only difference between frames was their relative velocity. And it does not make sense to give an abstract frame another quantity. But that changes in particle physics. Each massive particle that we see in the laboratory can be understood as the anchor of its own frame: the rest frame of the particle. For photons, the particles of light, we cannot find a rest frame as they are moving with the speed of light and the Lorentz transformation is only well defined for relative speeds that are smaller than the speed of light, as can be seen from eq. (90).

When we deal with particles it turns out that it is more convenient to use energy and momentum to describe a particle. From classical mechanics we know, that a particle that does not move also has no momentum. But the energy is more tricky. In classical mechanics there is no energy attributed to the mass of a particle. That changes with special relativity. Here we have Einsteins famous formula

$$m = \frac{E}{c^2} \ , \tag{95}$$

that relates the gravitational attraction with the energy content of the attracted object. This relation can also be used to assign the particle an energy, that corresponds to its mass. Although the term rest mass is redundant[14], it emphasizes how we measure the mass of a particle: we go to its rest frame and measure the gravitational attraction with scales or the inertial resistance to acceleration. Both measurements give the same result[15].

Using the rest energy and the vanishing momentum in the rest frame of the massive particle with mass m, we can assign it a four vector

$$p^\mu_{\text{rest}} = \left(\frac{E}{c} = mc, 0, 0, 0\right) \qquad \text{or} \qquad p^\mu_{\text{rest}} = \begin{pmatrix} \frac{E}{c} = mc \\ 0 \\ 0 \\ 0 \end{pmatrix} \ , \tag{96}$$

similar to the four vectors of eq. (56). We have to divide the energy by c in order to have the same units for all components of the four vector. Comparing the four momentum with the four vector of eq. (56) we can identify the energy with the time component and the three dimensional momentum with the space components.

When we see this particle of mass m moving in our frame with a velocity $\vec{v}$, we can make a Lorentz transformation using the velocity $\vec{v}$ to transform into the particles rest frame. That means we can do the Lorentz transformation with $-\vec{v}$ from the particles rest frame into our frame, where we will see the particle moving with the velocity $\vec{v}$. For simplicity we will align the $\hat{x}$-axis of our coordinate system again with the velocity: $\vec{v} = (v, 0, 0)$. Then the four momentum in our frame is

$$p^\mu = \Lambda(-v)^\mu{}_\nu \, p^\nu_{\text{rest}} \ , \tag{97}$$

[14] Mass alone means already the mass of the object when the object is at rest. In that sense the term rest mass is redundant. The old term *relativistic mass* for the energy of a moving object is unnecessary and misleading and I strongly suggest to forget and bury it as soon as possible. In that respect I go further than Griffiths [5], who already in 1987 tried to dissuade people to use it.

[15] Einsteins equivalence principle states, that the gravitational mass should be equal to the inertial mass. Several experiments measured the difference between the gravitational mass m_g and the inertial mass m_I and established an upper limit for the relative difference of $|m_g - m_I| < 10^{-12}(m_g + m_I)$.

or explicitly

$$\begin{pmatrix} \frac{E}{c} \\ p \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} mc \\ 0 \\ 0 \\ 0 \end{pmatrix} = \gamma mc \begin{pmatrix} 1 \\ \beta \\ 0 \\ 0 \end{pmatrix} \quad , \tag{98}$$

like the definition for the relativistic momentum used in [5]. That means we have now energy and momentum of a particle given by

$$E = \gamma mc^2 \qquad \text{and} \qquad \vec{p} = \gamma m\vec{v} \ . \tag{99}$$

We can use this relation also to obtain the velocity of a particle, if we have its energy and its momentum:

$$\vec{v} = \frac{\vec{p}c^2}{E} \ . \tag{100}$$

When we treat the four momentum as the quantity that describes the particle, we can ask, what the invariant of this four momentum is. The answer should not surprise us anymore: we can calculate this invariant in any frame and it becomes particularly simple in the rest frame of the particle as

$$p^2 = g_{\mu\nu}p^\mu p^\nu = \frac{E^2}{c^2} - (\vec{p})^2 = m^2c^2 - 0 = m^2c^2 \ . \tag{101}$$

So we can calculate the energy of the moving particle not only using the velocity, like in eq. (99), but also directly using only the mass and the momentum:

$$E^2 = m^2c^4 + (\vec{p})^2c^2 \ . \tag{102}$$

This is the most important form of relating mass, energy, and momentum in particle physics.

References

[1] Lecture notes by David Hogg:
`http://cosmo.nyu.edu/hogg/sr/sr.pdf`

[2] Chris Doran and Anthony Lasenby,
Geometric Algebra for Physicists
Cambridge University Press; ISBN-13: 9780521715959 (2003)

[3] Joseph A. Minahan,
Course Notes for Mathematical Methods of Physics
`http://www.teorfys.uu.se/people/minahan/Courses/Mathmeth/notes_v3.pdf`

[4] J.H. Heinbockel
Introduction to Tensor Calculus and Continuum Mechanics
Trafford Publishing, Canada; ISBN-13: 978-1553691334 (2001)
`http://www.math.odu.edu/~jhh/counter2.html`

[5] David Griffiths,
Introduction to Elementary Particles
John Wiley & Sons, Inc.; ISBN 0-471-60386-4 (1987)